Choice Models for Predicting Divisional Winners in Major League Baseball

DANIEL BARRY and J. A. HARTIGAN*

Major league baseball in the United States is divided into two leagues and four divisions. Each team plays 162 games against teams in the same league. The winner in each division is the team winning the most games of the teams in that division. We wish to predict the division winners based on games played up to any specified time. We use a generalized choice model for the probability of a team winning a particular game that allows for different strengths for each team, different home advantages, and strengths varying randomly with time. Future strengths and the outcomes of future games are simulated using Markov chain sampling. The probability of a particular team winning the division is then estimated by counting the proportion of simulated seasons in which it wins the most games. The method is applied to the 1991 National League season.

KEY WORDS: Baseball; Change point models; Markov chain Monte Carlo.

1. INTRODUCTION

Major league baseball is played by 26 teams divided into two leagues, the American League and the National League. The American Leage (AL) has two divisions, the AL East and the AL West, each consisting of 7 teams. The National League (NL) has two divisions, the NL East and the NL West, each also consisting of 7 teams. Each of the 26 teams plays 162 games in a season against teams in their own league. At the end of the regular season, the teams with the best record in each division play one another in a best-of-7-game series to decide the league champions. The two league champions then meet in a best-of-7-game series called the World Series.

The data for a particular season is available from the weekly magazine *The Sporting News,* which publishes a schedule of all games in early March and publishes the results of all games from the previous week during the season. Summary data for the 1991 National League season appears in Table 1. The table presents the results of home and away games for each pair of teams in four 6-week periods.

We will estimate the probability of winning their division for each of the teams, given the results of all games played up to a certain date and the list of remaining games for each team. We wish to allow for differing strengths for each team, for differing home advantages, and for changing strengths over time. It will be seen from Table 1 for example, that Atlanta had a poor record before the All-Star break (39 of 79) and a good record after the break (55 of 83).

We use a choice model for predicting the outcome of each game. The parameters of the model depend on the teams involved, on which team is playing at home, and on time. Markov chain sampling is used to simulate the outcomes of future games and so predict the eventual division winners. We argue that it is necessary to allow for changing team strengths with time, because some teams appear to change noticeably over a season. We present a variety of predictions of outcomes for the 1991 National League season. One prediction, at a certain point in the season when Atlanta was 2 games behind the Dodgers in number of wins, was that Atlanta had a better chance of winning the division than did the Dodgers. This occurred because Atlanta appeared stronger in the second half of the season and that strength was projected to the remaining games. In general, allowing for changing strengths encourages more conservative probability estimates; that is, the team with the best record has a lower probability of winning under the changing strength model than under a fixed strength model, because it is likely to have lower future strength than its record indicates.

Of course, other variables influence the probability of winning. In baseball, the starting pitcher has a substantial effect on the final outcome of each game. We judged that this effect would average out in a relatively short period. Starting pitchers work in a regular rotation, so that the total number of wins in a couple of weeks could be predicted from an estimate of the *average* ability of the team's starting pitchers. Collecting and analyzing the data for the rather large number of starting pitchers would be a formidable task. We suspect that including this variable would change the estimates for particular games quite a bit, and so the predictions as the season draws to a close (with 4 or 5 games remaining for each team) would be substantially affected. Toward the end of the season, you need to know who is pitching.

Point spreads—the differences in scores between the contending teams—were examined for the National Football League by Harville (1980). Harville predicted future point spreads using past point spreads observed over several seasons. The expected point spread in any game is the difference between a parameter for the home team and one for the away team; these parameters are the same in each year and vary from year to year according to an autoregressive process with lag 1. Thus Harville's model allows for differences between teams, for home field advantage (the same for all teams), and for change over time. The parameters are estimated by least squares and likelihood techniques. One outcome is an estimated distribution for the score difference in a particular future game, which Harville compared to the point spread offered by professional gamblers.

There is little in the statistical literature on predicting season outcomes in baseball or any other sport. Mosteller (1952)

* Daniel Barry is Lecturer in Statistics, University College, Cork, Ireland. J. A. Hartigan is Professor of Statistics, Yale University, New Haven, CT 06520. This research was partially supported by National Science Foundation Grant DMS-8617919.

Journal of the American Statistical Association
September 1993, Vol. 88, No. 423, Applications and Case Studies

Table 1. Win Records in the National League, by Time Period, Team Played, and Home Advantage (each pair of numbers is the number of home team wins, followed by the number of away team wins)

	Away team													
Home	*PI*	*NY*	*SL*	*CC*	*PH*	*MO*	*LA*	*AT*	*CR*	*SF*	*SD*	*HO*	*Totals*	
	Period 1: April 8 to May 26													
PI	00	21	12	31	30	21	00	12	30	00	00	12	16	9
NY	12	00	12	12	21	31	11	00	00	21	11	00	12	11
SL	00	00	00	21	31	21	00	21	21	00	00	21	13	6
CC	21	00	12	00	30	30	00	21	12	00	00	20	14	6
PH	00	12	21	21	00	00	12	00	00	20	31	00	11	7
MO	03	12	21	00	12	00	20	00	00	30	12	00	10	10
LA	00	21	00	00	21	30	00	12	00	31	03	00	11	8
AT	20	00	20	21	00	00	02	00	12	12	13	12	10	12
CR	11	00	21	31	00	00	12	02	00	30	12	21	13	10
SF	00	12	00	00	12	13	12	00	00	00	12	30	8	11
SD	00	12	00	00	12	03	22	00	12	21	00	00	7	12
HO	03	00	12	12	00	00	22	12	21	21	21	00	11	14
Totals	610	810	1211	149	169	149	1013	710	108	186	1015	116		
	Period 2: May 27 to July 7													
PI	00	00	00	21	00	12	11	00	00	12	22	00	7	8
NY	00	00	00	00	12	30	00	12	21	00	00	12	8	7
SL	03	21	00	12	30	00	12	00	00	22	13	00	10	13
CC	22	03	21	00	00	00	13	00	00	21	12	00	8	12
PH	03	03	13	00	00	21	00	21	03	00	00	22	7	16
MO	03	04	00	03	00	00	00	30	12	00	00	30	7	12
LA	31	00	21	30	00	00	00	21	22	02	00	21	14	8
AT	00	22	00	00	21	22	13	00	12	00	00	00	8	10
CR	00	21	00	00	22	22	00	00	00	00	30	21	11	6
SF	21	00	03	40	00	00	00	22	21	00	21	03	12	11
SD	12	00	12	21	00	00	12	03	00	12	00	40	10	12
HO	00	22	00	00	21	12	00	12	21	00	00	00	8	8
Totals	815	816	610	127	106	119	511	1111	1012	69	98	149		
	Period 3: July 11 to August 25													
PI	00	30	31	00	21	00	00	21	21	21	00	21	16	6
NY	21	00	21	04	00	00	22	00	12	21	31	00	12	12
SL	31	30	00	00	00	11	30	12	21	00	00	30	16	5
CC	00	40	00	00	21	12	00	21	21	00	12	31	15	8
PH	30	00	00	30	00	40	30	00	00	22	20	00	17	2
MO	00	00	12	12	04	00	40	00	00	12	12	00	8	12
LA	00	21	00	00	30	21	00	00	21	00	30	22	14	5
AT	40	00	40	21	12	00	00	00	00	22	12	30	17	7
CR	04	00	03	20	00	00	13	13	00	22	00	00	6	15
SF	00	30	00	00	30	20	30	12	13	00	00	00	13	5
SD	00	21	00	00	12	21	00	22	21	00	00	21	11	8
HO	12	00	30	12	00	12	30	00	00	12	22	00	12	10
Totals	138	172	137	99	1210	137	195	911	1210	1212	139	155		
	Period 4: August 26 to October 6													
PI	00	21	20	20	21	21	13	00	00	00	11	00	12	7
NY	12	00	21	01	21	22	00	03	00	00	00	12	8	12
SL	11	21	00	21	11	22	00	00	00	20	02	00	10	8
CC	11	02	30	00	12	14	11	00	00	21	00	00	9	11
PH	12	21	11	21	00	02	00	21	12	00	00	11	10	11
MO	21	00	03	00	02	00	00	21	30	00	00	30	10	7
LA	02	00	21	30	00	00	00	21	20	21	21	20	15	6
AT	00	20	00	00	00	20	21	00	12	20	20	21	13	4
CR	00	03	00	00	20	20	11	03	00	11	12	21	9	11
SF	03	00	21	02	00	00	21	20	20	00	12	12	10	11
SD	12	00	30	21	00	00	11	02	30	30	00	11	14	7
HO	00	11	00	00	21	00	02	03	21	12	02	00	6	12
Totals	714	99	157	116	108	1111	810	814	145	135	710	138		

Note: Abbreviations for National League teams: *National League East:* PI Pittsburgh Pirates; NY New York Mets; SL St. Louis Cardinals; CC Chicago Cubs; PH Philadelphia Phillies; MO Montreal Expos. *National League West:* LA Los Angeles Dodgers; AT Atlanta Braves; CR Cincinnati Reds; SF San Francisco Giants; SD San Diego Padres; HO Houston Astros.

considered questions of home field advantage and serial dependence in the World Series. Smith (1956) proposed a method for adjusting baseball standings for the strength of teams played. Rubin (1958) presented an analysis of baseball scores by innings. There has been much written about salary evaluation for professional baseball players; for example, Lackritz (1990). Bennett and Flueck (1983) studied a player's value to his team.

The technique we develop should have application in various sports leagues, but especially in baseball and basketball, where each team plays many games in a precisely specified schedule. College basketball, with a huge number of teams only loosely connected by playing each other, would be especially challenging and interesting; it would be necessary with teams of diverse initial strengths to use the previous season's records to set the prior strengths. Football has relatively few games, and thus it makes less sense to try to estimate change in strength over time. Harville (1980) considered several years' records with change allowed between years. In hockey just about all the teams qualify for the playoffs, so the prediction of divisional winners is of less interest.

2. THE MODEL

A league consists of B baseball teams, each of which plays G games. The total number of games played is $T = BG/2$; it will be assumed that the games start at T different times, which will be indexed $1, 2, \ldots, T$. For the National League, the parameters are $B = 12$, $G = 162$, and $T = 972$.

At time t, a game is played between home team h_t and away team a_t. The random variable W_t is set equal to 1 if the home team wins and to 0 if the away team wins.

The probability model used is a *choice* model (Bradley and Terry 1952):

$$p_t = P\{W_t = 1 | H, s\} = \frac{H(h_t)S_t(h_t)}{H(h_t)S_t(h_t) + S_t(a_t)},$$

where $H(b)$ is the home field advantage for team b and $S_t(b)$ is the strength of team b at time t.

We assume that the wins $W_1, \ldots, W_T$ are independent given the parameters H and S.

We will compare the full model with a number of specializations of this model, in some of which the home field advantage is constant across teams, in some of which the team strength is constant over time, and in some of which the team strength is constant over teams. In the full model, there are a huge number of parameters when variation over time is allowed—many more parameters than the number of games. It is, therefore, necessary to strongly control variation of strength over time.

We specify a prior distribution on the strengths S that allow $S_t(b)$ to change slowly with time. The strength sequences S for different teams are assumed independent:

a. $S_1(b)$ is uniform over $\ldots \lambda^{-2}, \lambda^{-1}, 1, \lambda, \lambda^2, \ldots$
b. if b does not play at time t, $S_t(b) = S_{t-1}(b)$,
c. if b does play at time t, and $t > 1$, $S_t(b) = S_{t-1}(b)/\lambda$ with probability $\alpha/2$, $S_t(b) = S_{t-1}(b)$ with probability $1 - \alpha$, $S_t(b) = S_{t-1}(b)\lambda$ with probability $\alpha/2$.

The home advantage parameters $H(1), \ldots, H(B)$ are assumed independent of the strength S, and each is taken to have prior distribution uniform over $\{\ldots \lambda^{-2}, \lambda^{-1}, 1, \lambda, \lambda^2, \ldots\}$. The parameter λ is the proportional change in strength. We select λ (in Sec. 4) so that there are not too many changes in the strengths over the season. The parameter α is the probability of change; it is taken to be uniformly distributed between 0 and 1. For a given λ, the parameter α will determine how many changes take place in the sequence of team strengths. If λ is set large, then α will be estimated from the data to be small, because just a few changes with big jumps will be sufficient to model the actual win sequences. If λ is set small, then α will be estimated large, because now many changes are necessary. The final predictions of the winners are not much affected, because in either case the set of possible strength sequences for each team are not much different. But the computations are roughly proportional to the number of changes in all a team's strengths, so there is an interest in choosing λ as large as possible while still being able to identify modest changes in a team's performance.

Our aim is to predict the team with the most wins at the end of the season, based on the record to date at time t. This requires predicting the outcome of each game at times $t + 1, \ldots, T$, using the record of games at times $1, \ldots, t$. We need to compute the probability distribution for $W_{t+1}, \ldots, W_T$ given $W, \ldots, W_T$. This is done using Equation (4), which simplifies computation by conditioning on the unknown home advantage and strength parameters H, S; the future record $W_{t+1}, \ldots, W_T$ is independent of the past record $W_1, \ldots, W_t$ given H and S.

In (1) compute the probability of the past observations $W_1, \ldots, W_t$ given H, S; in (2) compute the posterior probability of H, S given $W_1, \ldots, W_t$; and in (3) compute the probability of the future observations $W_{t+1}, \ldots, W_T$ given H, S. Finally in (4) sum over the products of the last two quantities to obtain the desired conditional probability of the future given the past:

$$p(W_1, \ldots, W_t | H, S) = \prod_{r=1}^{t} \frac{[H(h_r)S_r(h_r)]^{W_r}(S_r(a_r))^{1-W_r}}{H(h_r)S_r(h_r) + S_r(a_r)}, \quad (1)$$

$$p(H, S | W_1, \ldots, W_t) = p(H, S)p(W_1, \ldots, W_t | H, S)/p(W_1, \ldots, W_t), \quad (2)$$

$$p(W_{t+1}, \ldots, W_T | H, S) = \prod_{r=t+1}^{T} \frac{[H(h_r)S_r(h_r)]^{W_r}S_r(a_r)^{1-W_r}}{H(h_r)S_r(h_r) + S_r(a_r)}, \quad (3)$$

$$p(W_{t+1}, \ldots, W_T | W_1, \ldots, W_t) = \sum_{H,S} p(W_{t+1}, \ldots, W_T | H, S)p(H, S | W_1, \ldots, W_t). \quad (4)$$

It is not practicable to explicitly sum over all possible H, S. Instead, the sum is estimated by simulation. The general plan for the simulation is as follows: at time t, data is available in the form $W_1, \ldots, W_t$. We develop the sampling scheme again using the fact that the observations $W_1, \ldots, W_T$ are independent given H, S.

Thus using Equation (2), generate a random sample from

$$H, S \quad \text{given} \quad W_1, \ldots, W_t.$$

Using Equation (3), generate a random sample from

$$W_{t+1}, \ldots, W_T \quad \text{given} \quad H, S.$$

The sample $W_{t+1}, \ldots, W_T$ is now a sample from the posterior distribution of $W_{t+1}, \ldots, W_T$ given $W_1, \ldots, W_t$, according to Equation (4).

The outcome $W_{t+1}, \ldots, W_T$ together with the given $W_1, \ldots, W_t$ determines the winning team. The probability of winning the division is estimated by counting over all simulations, the number of times the team has the most wins of teams in its division. (In the case of ties, there will be a playoff; we model each playoff game to give each team probability $\frac{1}{2}$ of winning. This permits us to settle an end-of-season tie for a particular team by counting it as a half a win for that team. L-fold ties would be counted as $1/L$ of a win for each team in the tie.)

Details of the simulation of H, S, given $W_1, \ldots, W_t$, using Markov chain sampling, are given in Appendix A.

3. IF STRENGTH DOES NOT CHANGE OVER TIME

It is prudent to ask whether the complications of dealing with time-changing strength are necessary. We considered some models with constant strengths over time:

$$H_1\text{: } S(b) = 1, \qquad H(b) = 1,$$

$$H_2\text{: } S(b) = 1, \qquad H(b) \text{ constant},$$

$$H_3\text{: } S(b) = 1, \qquad H(b) \text{ variable},$$

$$H_4\text{: } S(b) \text{ variable}, \qquad H(b) = 1,$$

$$H_5\text{: } S(b) \text{ variable}, \qquad H(b) \text{ constant},$$

and

$$H_6\text{: } S(b) \text{ variable}, \qquad H(b) \text{ variable}.$$

Table 2 provides the maximum likelihood estimates of the parameters, with their standard errors, based on the whole season data, for these models. (The parameters are not restricted to the grid used in the time-change model.) There is some evidence, based on likelihood ratio tests, that Model 5 with different team strengths and constant home advantage is adequate among the constant time-strength models; that is, there is a home advantage, but the home advantage is not noticeably different in the different teams. According to this model the home advantage is 1.20, which means that the home team has 1.20^2 times the odds of winning than it would have when playing the same team away. Though the extra complication of Model 6, which allows differing team strengths and home advantages, is not supported by the likelihood ratio tests, some of the home advantages estimated by Model 6 are sufficiently large to warrant further examination. One of the referees remarked that "Model 5 suggests

Table 2. Maximum Likelihood When Strength Does Not Change Over Time

	Maximum likelihood	Degrees of freedom	Twice 2	Log 3	Likelihood 4	Differences 5	From 6
1	−672	0	8	24	24	32	40
2	−668	1		16		24	32
3	−660	12					16
4	−660	11				8	16
5	−656	12					8
6	−652	23					

The maximum likelihood estimates of the H, S parameters:

	Actual percent wins	Home = 1; team = 1 H	Home = 1; team = 1 S	Home constant team = 1 H	Home constant team = 1 S	Home varies team = 1 H	Home varies team = 1 S	Home = 1; team varies H	Home = 1; team varies S	Home constant team varies H	Home constant team varies S	Home varies team varies H	Home varies team varies S
NL East													
PI	.60	1.0	1.0	1.2	1.0	1.7	1.0	1.0	1.4	1.2	1.4	1.0	1.5
SL	.52	1.0	1.0	1.2	1.0	1.5	1.0	1.0	1.0	1.2	1.0	1.8	.8
PH	.48	1.0	1.0	1.2	1.0	1.2	1.0	1.0	.9	1.2	.9	1.6	.7
CC	.48	1.0	1.0	1.2	1.0	1.2	1.0	1.0	.9	1.2	.9	1.5	.7
NY	.48	1.0	1.0	1.2	1.0	1.0	1.0	1.0	.9	1.2	.9	.9	1.0
MO	.44	1.0	1.0	1.2	1.0	.9	1.0	1.0	.8	1.2	.8	1.0	.8
NL West													
AT	.58	1.0	1.0	1.2	1.0	1.4	1.0	1.0	1.3	1.2	1.3	.9	1.4
LA	.57	1.0	1.0	1.2	1.0	2.0	1.0	1.0	1.3	1.2	1.3	1.9	1.0
SD	.52	1.0	1.0	1.2	1.0	1.1	1.0	1.0	1.0	1.2	1.0	.8	1.2
SF	.46	1.0	1.0	1.2	1.0	1.1	1.0	1.0	.9	1.2	.9	1.5	.7
CR	.46	1.0	1.0	1.2	1.0	.9	1.0	1.0	.8	1.2	.8	1.0	.8
HO	.40	1.0	1.0	1.2	1.0	.8	1.0	1.0	.7	1.2	.7	1.3	.6
Standard error				.1	0.0	.3	.0	.0	.2	.1	.2	.4	.3

NOTE: H_1—all parameters equal to 1; H_2—one home advantage parameter; H_3—team strengths equal to 1, different home advantages; H_4—different team strengths, home advantage equal to 1; H_5—different team strengths, constant home advantage; H_6—different team strengths and home advantages.

Table 3. Probability of Winning the Division Under Various Models, Using Records up to August 1, 1991

	Actual percent wins	Games played	H_1 Home = 1 team = 1	H_2 Home const team = 1	H_3 Home varies team = 1	H_4 Home = 1 team varies	H_5 Home const team varies	H_6 Home varies team varies	Changing strength over time
					East				
PI	61	99	74	74	83	93	93	94	74
NY	55	100	18	18	11	5	6	5	16
SL	53	100	8	7	8	1	0	1	8
CC	48	100	0	0	1	0	0	0	2
MO	43	100	0	0	0	0	0	0	0
PH	42	100	0	0	0	0	0	0	0
					West				
LA	58	100	70	71	84	81	82	77	55
AT	54	99	20	19	12	17	17	21	34
CR	50	98	6	6	2	1	1	1	4
SF	48	99	3	3	2	0	0	0	6
SD	48	101	1	1	0	0	0	0	1
HO	41	100	0	0	0	0	0	0	0

that Atlanta and Los Angeles are essentially equal in strength, while Model 6 reveals that Los Angeles is competitive thanks only to an overwhelming home field advantage; on a neutral field, Atlanta is nearly 50% stronger (1.4 to 1.0!) This is in contrast to the National League East, where Pittsburgh is unambiguously the dominant team regardless of model." The standard errors of the home advantage estimates in Model 6 are quite high, in part due to the dependence between estimates of home advantage and team strength. One would expect the home field advantage to be sustained from one season to the next, so that several seasons' data could sensibly be used for estimating these parameters.

We now consider using each of these models to predict the eventual winners of the two divisions, using data from games played up to the end of July. The predictions were computed by simulating the remaining games in the season 2,500 times, using parameters estimated from these data, and for each team calculating the proportion of times they won the most games in their division. Table 3 shows the predictions for the various models.

We see that the models that allow different strengths among the teams tend to be more confident in their predictions about the division winners; these models predict that the teams that have done well in the past will continue to do as well in the future and so give a high probability of a win to the team with the best record to date. Models that do not allow different strengths between teams, or allow the strength to change over time, are not so sure that the past record will extend to the future, and so are less confident that the team with the best record to date will continue to do well.

We cannot use standard likelihood methods to compare the model allowing changing strengths over time with the fixed strength models, because the very large number of discrete parameters in the changing strength models makes the usual χ^2 asymptotics nonapplicable. Our approach was to inhibit too-rapid change in the strengths through a prior probability model. In addition, maximum likelihood gives estimates of strengths for the season to date but does not determine future strengths, which are necessary to predict the rest of the season; the probability model for changing strengths is well set up to do that.

As evidence of changing strength over time, we compared for each team the proportions of home and away wins in each of the four periods of Table 1. Two chi-squared statistics testing equality of proportions are computed for each team—

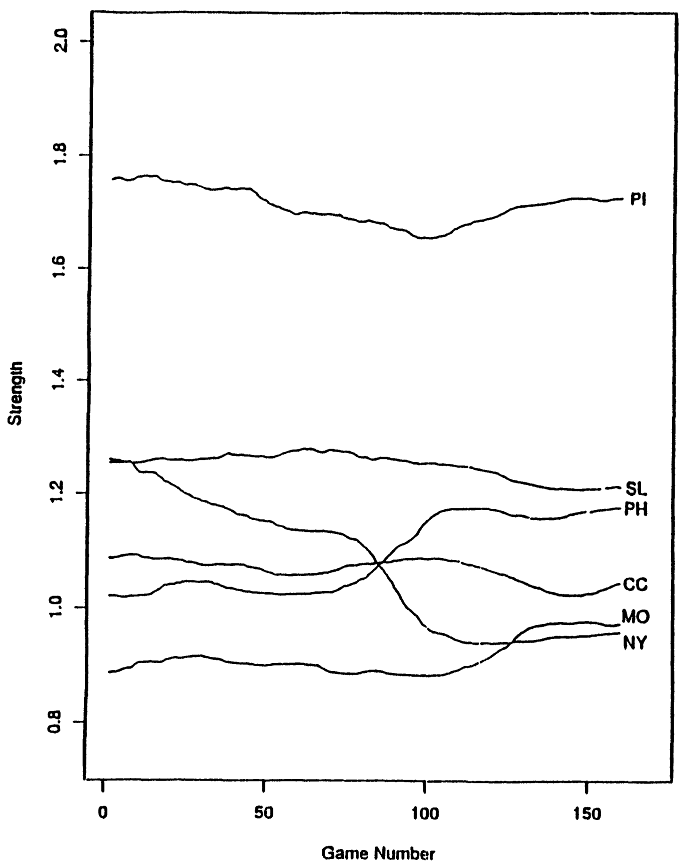

Figure 1. NL East Strengths.

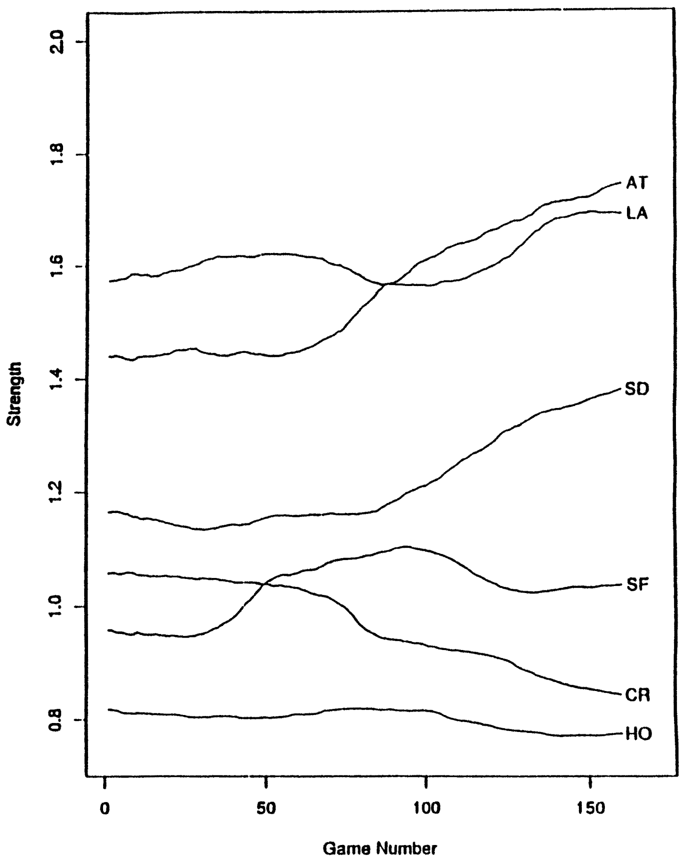

Figure 2. NL West Strengths.

one for home games and one for away games—and the statistics are summed over all teams. For instance, for the New York Mets the four home percentages are 52, 53, 50, and 40, with an overall percentage of 49; the four away percentages are 56, 67, 11, and 50, with an overall percentage of 47. This leads to a home χ^2 of .86 and an away χ^2 of 14.47. For all 12 teams, the total $\chi^2 = 108$ with nominal degrees of freedom 72. (There are dependencies between the numbers of wins of all teams and variations in the schedule between quarters, which make the usual χ^2 distribution inappropriate as a reference distribution.) To assess the significance of this value, we simulated the results for the entire season 10,000 times, taking the true probability model to be the maximum likelihood fit of Model 6, and calculated the test statistic each time. Only 59 of the simulated values were larger than the observed value of 108, which we take to be evidence of changing strength over time. (One referee advanced the alternative explanation that there might be dependencies within each period because teams tend to play sets of three or four games in a row against the same opponent. This particular source of dependency might be excluded by an examination of the results of all games between a particular pair of teams; in each set, the results should be binomial, and dependencies should increase the variance between sets. Other kinds of dependencies also might offer alternative explanations to changing strengths. Indeed, the time-change model is a way of absorbing time-local dependencies.)

Less formal evidence may be seen in Figure 1 and Figure 2, which show the posterior expected strengths at each time for each team, based on the complete season's data.

4. DATA ANALYSIS USING THE STRENGTH CHANGING OVER TIME

4.1 Markov Chain Sampling Implementation

The choice of which value of λ to use is somewhat arbitrary; we experimented with λ values in the range 1.01 to 2. The win probabilities for the various teams were not much affected. As a result, we selected $\lambda = 1.5$; the win probabilities $1/(1 + \lambda^k)$ then take possible values . . . .23, .32, .40, .50, .60, .69, .77, ... With this grid of possible probabilities, we need a fair bit of evidence before deciding that a team's win probability has increased from, say, .50 to .60, so that not too many changes in strength should occur. The change probability α and the amount of change λ serve a similar purpose; if λ is small, then the change probability α may be increased to cause the same net change over a long period. It is not necessary (and indeed not possible) to identify λ from the data. Over a long period of time T, the strength of a team will change by $\lambda^{Z\sqrt{\alpha T}}$, where Z is approximately unit normal; this means that we can estimate only $\lambda^{\sqrt{\alpha}}$.

In implementing the Markov chain sampling, we initialized α to .5 and H and S to 1. We began with 300 preparatory passes, each pass simulating a value of each parameter conditional on the rest. These passes were necessary to reach a point where the stationary distribution of the parameters was approximately the posterior distribution of the parameters given the observations to date. (We found that 1,000 or 5,000 preparatory passes did not much effect the final estimates.) The preparatory passes were followed by $M = 50$ n production passes, in which both parameters and future game outcomes are simulated. For each of the n sets of 50 passes we estimated the probability of each team winning its division as the number of simulated wins divided by 50. Our final estimate of the probability of winning is the average of these n estimates. The sequence of n estimates for each team allows us to estimate the first-order autocorrelation among the estimates and to obtain a standard error for the final

Table 4. Posterior Means of Home Advantage Parameters

	Parameter	Actual home percentage	Actual away percentage
		East	
PI	1.03	.63	.58
SL	1.97	.60	.43
PH	1.66	.56	.41
CC	1.76	.55	.40
NY	1.05	.49	.47
MO	1.01	.46	.42
		West	
AT	.95	.59	.57
LA	2.09	.67	.48
SD	.84	.52	.52
SF	1.52	.53	.40
CR	1.04	.48	.43
HO	1.40	.46	.35

Table 5. Actual Win Percents and Predicted Probabilities of Winning Using the Markov Model

	5/1		6/1		7/1		8/1		9/1		10/1	
					NL East							
PI	65	40	67	68	63	71	61	74	60	97	61	100
SL	62	17	52	6	54	16	53	8	54	3	52	0
NY	60	28	58	15	54	10	55	16	49	0	48	0
PH	43	5	47	2	43	1	42	0	48	0	48	0
CC	48	6	51	8	45	1	48	2	50	0	47	0
MO	35	4	43	1	43	1	43	0	41	0	45	0
					NL West							
LA	50	15	57	33	61	69	58	55	52	25	58	53
AT	44	6	57	46	51	5	54	34	56	74	57	47
SD	52	35	49	7	49	4	48	1	48	0	51	0
CR	58	32	50	12	55	19	50	4	50	1	47	0
SF	40	7	33	0	44	3	48	6	48	0	46	0
HO	42	5	38	2	39	0	41	0	41	0	40	0

estimate taking this autocorrelation into account. We chose a value for n so that the standard errors of our estimates were uniformly no larger than .01. The value of n required depended on how many games had already been played and how close the individual division races were. In the early part of the season when data were scarce, no clear winners could be predicted, and a large value of n was required. As the season progressed, large values of n were required at points where division races became close. For example, on May 1, $n = 210$ was required; on July 1, $n = 60$ was required; and on August 1, $n = 110$ was required.

4.2 The Results

We sampled the posterior distribution given the data for the entire season using Markov chain sampling with $n = 200$. The mean value for α was .006, and the mean number of changes in strength was 11 (about one per team per season). The posterior means for the home advantage parameters are shown in Table 4. The values are very similar to the Model 6 maximum likelihood estimates given in Table 2. Define the average strength of team b at time t by $A_t(b) = S_t(b)(1 + H(b))/2$, the average of the strengths at home and away. Figure 1 contains plots of the posterior means of average strength versus game number for each of the six teams in the NL East. Figure 2 shows similar plots for the NL West. These plots show the changing fortunes of the teams over the course of the season. Figure 1 reveals the consistent dominance of the Pittsburg Pirates, the consistent poor performance of the Montreal Expos, the sharp decline in the fortunes of the New York Mets after the midway point of the season, and the equally sharp improvement in the fortunes of Philadelphia towards the latter half of the season. The most interesting feature of Figure 2 is the way in which it reflects the close contest between Atlanta and the Los Angeles Dodgers, with Atlanta coming from a long way behind to edge out the Dodgers at season's end. In both plots, we see that many teams exhibited no marked changes in strength over the season. Thus although the model does allow for the possibility of change, it is not overly sensitive to short runs of good or bad performances.

We used the model for predicting the eventual winners of each of the two divisions at various times during the season. We made predictions using data from games played up to the beginning of each month from May to October; these are shown in Table 5. The predictions are more conservative than those based on models in which strength does not change; the changing strength model gives lower probabilities to the leading team. This was especially so in the case of the NL West, where the changing strength model gave much higher probability to Atlanta winning than did the constant models. The constant strength model gave equal weight to all games played to date in estimating the win probabilities for future games. The changing strength model was able to take into account the fact that Atlanta was steadily improving as the season went on and so based predictions on their most recent form.

APPENDIX A: MARKOV CHAIN SAMPLING

The Markov chain sampling technique originated in statistical physics (Hammersley and Handscomb 1964; Hastings 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953). It has become increasingly popular for simulating samples from complex multivariate densities, such as posterior densities that arise in Bayesian analysis (Gelfand and Smith 1990). Perhaps the most striking applications have been in image processing (Cross and Jain 1983; Geman and Geman 1984).

To sample from a finite random variable Z having density f, we devise a Markov chain with transition probabilities $f(Y\,|\,Z)$ chosen so that the corresponding stationary probabilities are $f(Z)$. We begin with some arbitrary starting value Z_0 and compute the Markov sequence $Z_1, Z_2, \ldots Z_m$ according to the transition probabilities $f(Y\,|\,Z)$. If the chain is irreducible, (i.e., every state can be reached from every other with positive probability), then the frequency of occurrence of the states Z_t converges to the stationary distribution $f(Z)$. We thus use the Z_i as our sample from the density f.

One choice of transition matrix is as follows: Assume that U and V are finite variables with joint density $f_{U,V}$. Let f_U and f_V be the corresponding marginal densities, and let $f_{U|V}$ and $f_{V|U}$ be the corresponding conditional densities.

We begin with some initial states U^0, V^0; a transition matrix for generating new states U^i, V^i from states U^{i-1}, V^{i-1} is defined as follows:

$$f(u^i, v^i\,|\,u^{i-1}, v^{i-1}) = f_{U|V}(u^i\,|\,v^{i-1}) f_{V|U}(v^i\,|\,u^i).$$

Thus we generate U^i from the conditional density of U given $V = v^{i-1}$ and then generate V^i from the conditional density of V given

$U = u^i$. The joint density $f_{U,V}$ is a stationary distribution for this transition matrix. To see this, suppose that the initial states $U^0, V^0 \sim f_{U,V}$. Then $V^0 \sim f_V$ and, because $U^1 \mid V^0 \sim f_{U|V}$, it follows that $U^0, V^1 \sim f_{U,V}$; repeating the argument, it follows that $U^1, V^1 \sim f_{U,V}$. Thus $f_{U,V}$ has been shown to be stationary for the given transition matrix. When the chain is irreducible, the limiting distribution of the sampled points U^i, V^i is just the required density $f_{U,V}$. And more generally, if we have n variables $U_1, \ldots, U_n$, we generate a sample with limiting distribution $f_{U_1 U_2 \ldots U_n}$ by beginning with some $u_1^0, \ldots, u_n^0$ and making the transitions one variable at a time conditional on the values of the other variables; we select U_k^i from the density $f(u_k \mid u_1^i, \ldots, u_{k-1}^i, u_{k+1}^{i-1}, \ldots, u_n^{i-1})$. This technique works best when the joint density $f(u_1, \ldots, u_n)$ is in simple form, for then all the conditional densities are likewise simple.

For the baseball problem, the individual conditional samples in the one-variable transitions require generating an integer random variable Z from a density $f(z)$. We use Markov chain sampling for this problem also, with transition probabilities

$$f(Y = i - 1 \mid Z = i) = \frac{f(i-1)}{2[f(i) + f(i-1)]}$$

$$f(Y = i \mid Z = i) = \frac{f(i)}{2[f(i) + f(i-1)]} + \frac{f(i)}{2[f(i) + f(i+1)]}$$

$$f(Y = i + 1 \mid Z = i) = \frac{f(i+1)}{2[f(i) + f(i+1)]}.$$

It may be shown that f is stationary for these transition probabilities. Ordinary sampling techniques would require that we compute f at all values in its range, whereas this technique requires computation at only three values.

APPENDIX B: MARKOV CHAIN SAMPLING TO PREDICT BASEBALL

Markov chain sampling is used to generate samples from the posterior distribution of H, S given $W_1, \ldots, W_t$. (The complete sampling procedure described in Sec. 2 also requires generating the future observations $W_{t+1}, \ldots, W_T$ given H, S, but this requires only simple independent Bernoulli random sampling.) Convergence to the stationary distribution is more rapid if the changes in S are simulated rather than S itself. (The reason is that the changes in S are approximately independently distributed, rather than the actual S values. If a team's strength jumps at a certain time, we expect the new value to be stable for a while; thus if a value of S is changed during the simulation, we would like the following values to be changed as well. Therefore we simulate changes in S, with most changes being 0.) Define the parameters Δ, δ by

$$\lambda^{\Delta_s(b)} = S_s(b)/S_{s-1}(b), \quad s > 1,$$

$$\lambda^{\Delta_1(b)} = S_1(b),$$

$$\lambda^{\delta(b)} = H(b).$$

The *strength change* parameter Δ consists of the integers $\Delta_s(b)$, $1 \le s \le T$, and $1 \le b \le B$. For $1 < s \le T$, the integers specify the change in strength for team b between time $s - 1$ and time s. Most of the $\Delta_s(b)$'s will be 0. Every so often, there is a $\Delta_s(b)$ of plus or minus 1, which increases or decreases the team's strength by λ. The $\Delta_1(b)$'s are integers determining the strength of team b at the beginning of the season, which will be estimated probabilistically by the results of the season.

The *home advantage* parameter δ consist of the integers $\delta(b)$, which determine the homefield advantage for team b.

The *change probability* parameter α is the prior probability of a change at time s.

The Markov chain sampling is done on the parameters Δ, δ, and α. It is sufficient to specify the joint posterior density $f(\Delta, \delta, \alpha)$; the Markov chain sampling proceeds by selecting each real valued parameter in turn and then sampling according to its posterior density given the data and all other parameters. The prior densities for $\Delta_1(b)$ and $\delta(b)$, $1 \le b \le B$ are assumed to be independent uniforms over the integers, corresponding to $S_1(b)$ and $H(b)$ being uniform over $\ldots \lambda^{-2}, \lambda^{-1}, 1, \lambda, \lambda^2, \ldots$.

We know the results of games up to time t. Let C denote the set of teams and times when the teams play and change might occur, $C = \{b, s \mid 1 < s \le t,\ h_s = b \text{ or } a_s = b\}$. The joint density of parameters and data is

$$\prod_C \left(\frac{\alpha}{2}\right)^{|\Delta_s(b)|} (1-\alpha)^{1-|\Delta_s(b)|} \prod_{s=1}^{t} \frac{\lambda^{(\sum_{r=1}^{s} \Delta_r(h_s) + \delta(h_s))W_s + \sum \Delta_r(a_s)(1-W_s)}}{\lambda^{\sum \Delta_r(h_s) + \delta(h_s)} + \lambda^{\sum \Delta_r(a_s)}}.$$

The sequence of conditional sampling is arranged as follows:

1. Sample the change probability α from a beta distribution with parameters $1 + \sum_C |\Delta_s(b)|$ and $1 + \sum_C (1 - |\Delta_s(b)|)$. Note that $\sum_C |\Delta_s(b)|$ is the total number of strength changes in the present set of parameters.

2. Sample the home advantage $\delta(b)$ from a conditional probability formed from games played by b at home,

$$\prod_{\{s \mid h_s = b\}} \frac{\lambda^{(\delta(b) + \sum_{r=1}^{s} \Delta_r(b))W_s + \sum \Delta_r(a_s)(1-W_s)}}{\lambda^{\delta(b) + \sum \Delta_r(b)} + \lambda^{\sum \Delta_r(a_s)}},$$

where $\delta(b)$ is an integer; Markov chain sampling moves $\delta(b)$ to $\delta(b) - 1$, $\delta(b)$ or $\delta(b) + 1$.

3. Sampling the beginning strength $\Delta_1(b)$ from a conditional probability formed from games played by b,

$$\prod_{\{s \mid h_s = b, s \le t\}} \frac{\lambda^{(\sum_{r=1}^{s} \Delta_r(b) + \delta(b))W_s + \sum \Delta_r(a_s)(1-W_s)}}{\lambda^{\sum \Delta_r(b) + \delta(b)} + \lambda^{\sum \Delta_r(a_s)}}$$

$$\times \prod_{\{s \mid a_s = b, s \le t\}} \frac{\lambda^{(\sum \Delta_r(h_s) + \delta(h_s))W_s + \sum \Delta_r(b)(1-W_s)}}{\lambda^{\sum \Delta_r(h_s) + \delta(h_s)} + \lambda^{\sum \Delta_r(b)}}.$$

4. Sample the past strength changes $\Delta_s(b)$, $2 \le s \le t$, from a conditional probability formed from games played by team b between times s and t:

$$\left(\frac{\alpha}{2}\right)^{|\Delta_s(b)|} (1-\alpha)^{1-|\Delta_s(b)|} \prod_{h_u = b, s \le u \le t} \frac{\lambda^{(\sum_{r=1}^{u} \Delta_r(b) + \delta(b))W_u + \sum \Delta_r(a_u)(1-W_u)}}{\lambda^{\sum \Delta_r(b) + \delta(b)} + \lambda^{\sum \Delta_r(a_u)}}$$

$$\times \prod_{a_u = b, s \le u \le t} \frac{\lambda^{(\sum_{r=1}^{u} \Delta_r(h_u) + \delta(h_u))W_u + \sum \Delta_r(b)(1-W_u)}}{\lambda^{\sum \Delta_r(h_u) + \delta(h_u)} + \lambda^{\sum \Delta_r(b)}}$$

(Results of games played before s are not relevant, because all strengths up to time s are included in the conditioning variables.)

5. Sample the future strength changes $\Delta_s(b)$, $s > t$, from

$$\left(\frac{\alpha}{2}\right)^{|\Delta_s(b)|} (1-\alpha)^{1-|\Delta_s(b)|}.$$

Note that the change may be plus or minus 1 with equal probability.

The main computational burden is in the repeated generation of the $\Delta_s(b)$; because $\Delta_s(b)$ takes values $-1, 0, 1$, we need only do three products for the computations. The products are over terms u, $s \le u \le t$, and these may be computed recursively as s increases.

[*Received January 1992. Revised December 1992.*]

REFERENCES

Bennett, J. M., and Flueck, J. A. (1983), "An Evaluation of Major League Baseball Offensive Performance Models," *The American Statistician*, 37, 76–82.

Bradley, R. A., and Terry, M. E. (1952), "Rank Analysis of Incomplete Block Designs I: The Method of Paired Comparisons," *Biometrika*, 39, 324–345.

Cross, G. C., and Jain, A. K. (1983), "Markov Random Field Texture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* PAMI-5, 25-39.

Gelfand, A. E., and Smith, F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association,* 85, 398-409.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* PAMI-6, 721-741.

Hammersley, J. M., and Handscomb, D. C. (1964), *Monte Carlo Methods,* London: Methuen.

Harville, D. A. (1980), "Predictions for National Football League Games via Linear-Model Methodology," *Journal of the American Statistical Association,* 75, 516-524.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika,* 57, 97-109.

Lackritz, J. R. (1990), "Salary Evaluation for Professional Baseball Players," *The American Statistician,* 44, 4-8.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics,* 21, 1087-1091.

Mosteller, F. (1952), "The World Series Competition," *Journal of the American Statistical Association,* 47, 355-380.

Rubin, E. (1958), "An Analysis of Baseball Scores by Innings," *The American Statistician,* 12, 21-22.

Smith, J. H. (1956), "Adjusting Baseball Standings For the Strength of Teams Played," *The American Statistician,* 10, 23-24.